TILBURG ◆ UNIVERSITY

Tilburg University

Tail processes under heavy random censorship with applications

Einmahl, J.H.J.; Ruymgaart, F.H.

Published in:
Journal of Nonparametric Statistics

Publication date:
1995

Citation for published version (APA):
Einmahl, J. H. J., & Ruymgaart, F. H. (1995). Tail processes under heavy random censorship with applications. *Journal of Nonparametric Statistics*, *5*(3), 223-235.

Nonparametric Statistics, Vol. 5, pp. 223–235
Reprints available directly from the publisher
Photocopying permitted by license only

TAIL PROCESSES UNDER HEAVY RANDOM CENSORSHIP WITH APPLICATIONS

JOHN H. J. EINMAHL[1] and FRITS H. RUYMGAART[2]

[1]*Eindhoven University of Technology*
[2]*Texas Tech University, Lubbock*

(Received: July 28, 1993; Revised: March 2, 1995; Accepted: March 21, 1995)

We consider a type of heavy random censoring where the number of uncensored observations still tends to infinity. Under natural conditions the life distribution can be locally analyzed by generalizing tail empirical processes to the heavily censored case. A uniform central limit theorem for the tail product-limit process and the tail empirical cumulative hazard process is established. Statistical applications include a local confidence band for the cumulative life distribution and a test concerning the value of its density at the origin.

KEYWORDS: Heavy censoring, tail product-limit and tail empirical cumulative hazard process, uniform central limit theorem.

1. INTRODUCTION AND PRELIMINARIES

Most extensively studied in the random censoring literature is the situation where the distribution of the censoring variables is fixed. A fundamental result on the weak convergence of empirical processes under fixed censoring was first obtained in Breslow and Crowley (1974). Besides this standard case extremely heavy censoring is considered in Wellner (1985) where the distribution of the censoring variables tends to the degenerate distribution at 0 at such a high rate that the number of uncensored observations remains bounded for large sample sizes so that the usual asymptotics cannot be performed. In this note we consider a type of censoring that will still be called "heavy" in the sense that the censoring distribution is again degenerate at 0 in the limit but at a sufficiently slow rate to ensure that the number of uncensored observations tends to infinity so that asymptotic considerations remain possible. Mathematically, such heavy censoring requires asymptotics different from the asymptotics for the ordinary censoring in much the same way as the binomial distribution requires different asymptotics when the success probability is small. In practice, however, one is almost always given a fixed number of data points and the usual way to asymptotic considerations is by embedding the actual censoring problem in a sequence of such problems where, in some suitable manner, the censoring distribution converges to the distribution degenerate at 0. What "suitable" means is often dictated by mathematical convenience or tractability.

In order to specify an embedding which conveniently leads to alternative asymptotic results for our heavy censoring problem we first need to introduce some notation. For each $n \in \mathbb{N}$ let $(X_1, Y_{n1}), \ldots, (X_n, Y_{nn})$ be independent and identically distributed bivariate random vectors with $X_i \perp\!\!\!\perp Y_{ni}$ for each $i = 1, \ldots, n$. We observe the random variables $Z_{ni} = X_i Y_{ni}$ (i.e., the minimum of the variable X_i of interest and the censoring variable Y_{ni}) as well as the indicators $\delta_{ni} = \mathbb{1}\ \{X_i \leqslant Y_{ni}\}$ (which assume the value 0 if and only if censoring actually takes place). Since the X_i represent the life time they are nonnegative so that their common cumulative distribution function (c.d.f.) F has support in $[0, \infty)$. This c.d.f., which does *not* depend on n, is supposed to be at least continuous. The censoring variables Y_{ni} are also nonnegative but for the purpose of our embedding their common c.d.f. G_n *does* depend on n.

Our first assumption now is that for some continuous c.d.f. G we have

$$G_n(s) = G(s/a_n), \quad \text{for} \quad 0 \leqslant s \leqslant a_n, \quad G(1) < 1, \tag{1.1}$$

where the a_n, $n \in \mathbb{N}$, are strictly positive numbers satisfying

$$a_n \to 0, \quad \text{and} \quad nF(a_n) \to \infty, \quad \text{as } n \to \infty. \tag{1.2}$$

Note that the first part of condition (1.1) is equivalent to the Y_{ni}/a_n having c.d.f. G. The first part of conditions (1.1) and (1.2) entail heavy censoring since, indeed, we have $\mathbb{P}\{X_i \textit{ is censored at stage } n\} = \mathbb{P}\{X_i > Y_{ni}\} = \int_0^\infty \{1 - F(s)\} dG(s/a_n) = \int_0^\infty \{1 - F(a_n t)\} dG(t) \to 1$, as $n \to \infty$, by the dominated convergence theorem. The second condition in (1.2) prevents this convergence from being too fast. The next condition, more technical but rather natural in this context, is that F be regularly varying at 0, meaning that

$$\lim_{x \downarrow 0} \frac{F(xt)}{F(x)} = \tilde{F}(t), \quad t \geqslant 0. \tag{1.3}$$

In addition we want $\tilde{F}$ to be a c.d.f. on $[0, 1]$ which necessarily entails that

$$\tilde{F}(t) = t^\gamma, \quad 0 \leqslant t \leqslant 1, \quad \text{for some } \gamma > 0. \tag{1.4}$$

These conditions, in conjunction with (1.1) and (1.2), allow for a convenient time transformation defined in (2.11).

In Section 3 we will return to these conditions and see that a_n, G, and γ can be estimated from the data. This is important for practical application of our results.

Our model is completely different from Wellner's (1985) heavy censoring model. In our model we still have convergence of normal type, whereas in Wellner's model Poisson convergence takes place. This is implied by the fact that Wellner's model is inspired by the domain of attraction condition in extreme value theory, whereas our process ξ_n, defined in Section 2, is the natural extension of the tail empirical process to the random censorship model. However, it is easy to adapt Wellner's model in such a way that the Poisson convergence is replaced by convergence to a Gaussian limiting process. But even this adapted model is rather different from ours in various respects. A more detailed discussion of these differences would be rather technical and is therefore omitted.

A practical application of our heavy censoring model is, e.g., provided by airplane industry. Consider a type of airplane which has been in production for only a few years. For each of the first n produced airplanes we observe the number of flight hours up till now (Y_{ni}) or, if this happens earlier, the number of flight hours until a serious crack appears in the body of the airplane (X_i). Of course, here the variables of interest for the airplane factory are the X_i and, if the type of airplane under investigation is of good quality, it is obvious that heavy censoring occurs in the above situation.

The main theoretical tool is contained in the asymptotic behavior of the product-limit estimator of F in a neighborhood of the origin. This tool is presented in Section 2; its proof is deferred to the Appendix. The result boils down to weak convergence of the tail empirical process (see, e.g., Einmahl (1992)) generalized so as to allow for censored observations, and might be of independent interest. Statistical applications are considered in Section 3. These applications yield alternative approximate level-$(1-\alpha)$ local (near zero) confidence bands for F and level-α tests for the value of the density of F at 0. In Section 4 we briefly comment on these results.

2. WEAK CONVERGENCE OF CENSORED TAIL PROCESSES

We need to start this section with a short review of some basic concepts and relations and some more notation for which the reader is also referred to Shorack and Wellner (1986). Of fundamental importance are the functions

$$H_n(t) = \mathbb{P}\{Z_{ni} \leqslant t\}, \quad H_n^*(t) = \mathbb{P}\{Z_{ni} \leqslant t, \delta_{ni} = 1\}, \tag{2.1}$$

their empirical counterparts

$$\hat{H}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[0,t]}(Z_{ni}), \quad \hat{H}_n^*(t) = \frac{1}{n}\sum_{i=1}^{n} \delta_{ni}\mathbb{1}_{[0,t]}(Z_{ni}), \tag{2.2}$$

and the empirical processes

$$U_n(t) = \sqrt{n}(\hat{H}_n(t) - H_n(t)), \quad U_n^*(t) = \sqrt{n}(\hat{H}_n^*(t) - H_n^*(t)), \tag{2.3}$$

all defined for $t \geqslant 0$. It should be noted that $\hat{H}_n$ and $\hat{H}_n^*$ are unbiased estimators of H_n and H_n^* respectively.

We are interested in estimating F. Due to the heavy censoring we are forced to restrict ourselves to the left-hand tail. For the estimation relations between F and the cumulative hazard function

$$\Lambda(t) = \int_0^t \frac{1}{1 - F^-(s)}\, dF(s), \quad t \geqslant 0, \tag{2.4}$$

will be exploited, where for any right-continuous function with left-hand limits ψ: $[0, \infty) \to \mathbb{R}$ we write ψ^- for the left-continuous version. The cumulative hazard function is also of importance in its owns right. Since

$$\Lambda(t) = \int_0^t \frac{1}{1 - H_n^-(s)} \, dH_n^*(s), \quad t \geqslant 0, \tag{2.5}$$

a natural estimator, called the Nelson–Aalen estimator, appears to be

$$\hat{\Lambda}_n(t) = \int_0^t \frac{1}{1 - \hat{H}_n^-(s)} \, d\hat{H}_n^*(s), \quad t \geqslant 0. \tag{2.6}$$

The relation between F and Λ requires product-integration. A general introduction to this topic with application in censoring can be found in Gill and Johansen (1990), where it is in particular shown that

$$F(t) = 1 - \prod_0^t \{1 - d\Lambda(s)\}, \quad t \geqslant 0, \tag{2.7}$$

where $\prod$ denotes the product-integral. (By convention in expressions like (2.4)–(2.7) the interval of integration is assumed to be $(0, t]$.) Substituting the estimator $\hat{\Lambda}_n$, given in (2.6), in the expression on the right in (2.7) yields an estimator for F which is easily seen to be equal to

$$\hat{F}_n(t) = 1 - \prod_{i: \tilde{Z}_{ni} \leqslant t} \left(1 - \frac{1}{n - i + 1}\right)^{\tilde{\delta}_{ni}}, \quad t \geqslant 0, \tag{2.8}$$

where the $\tilde{Z}_{ni}$ denote the ordered Z_{ni} and $\tilde{\delta}_{ni}$ the corresponding δ_{ni}. This estimator is the well-known Kaplan–Meier or product-limit estimator.

We are now ready to introduce the tail product-limit process

$$\xi_n(t) = \sqrt{\frac{n}{F(a_n)}} \left(\hat{F}_n(ta_n) - F(ta_n)\right), \quad 0 \leqslant t \leqslant 1, \tag{2.9}$$

and the tail empirical cumulative hazard process

$$\beta_n(t) = \sqrt{\frac{n}{F(a_n)}} \left(\hat{\Lambda}_n(ta_n) - \Lambda(ta_n)\right), \quad 0 \leqslant t \leqslant 1. \tag{2.10}$$

Furthermore let

$$D(t) = \int_0^t \frac{1}{1 - G(s)} \, ds^\gamma, \quad 0 \leqslant t \leqslant 1, \tag{2.11}$$

denote a suitable time transformation, well defined because $G(1) < 1$, and W a standard Wiener process. Although the weak convergence of the processes β_n, established below, is a basic ingredient for statistical inference about the cumulative hazard function and the hazard rate, here we will only use the result to conveniently deal with the related but more complicated processes ξ_n. For statistical application we restrict ourselves to those related to the latter processes.

THEOREM 2.1 Under assumptions (1.1)–(1.4) there exists a special construction of the processes ξ_n, β_n ($n \in N$) and W, defined on one and the same probability space such that

$$\sup_{0 \leqslant t \leqslant 1} |\xi_n(t) - W \circ D(t)| \to_p 0, \quad \text{as } n \to \infty, \tag{2.12}$$

$$\sup_{0 \leqslant t \leqslant 1} |\beta_n(t) - W \circ D(t)| \to_p 0, \quad \text{as } n \to \infty. \tag{2.13}$$

3. A CONFIDENCE BAND AND A TEST

Although our final results will be presented for unknown a_n which will be estimated from the data, the first part of this section is based on the assumption that the a_n are known in order to avoid undue technicalities.

For the construction of a confidence band for F near 0 it turns out that we need to estimate the value of the function D in (2.11) at the point $t = 1$. For this purpose we estimate the c.d.f. $G(s) = G_n(sa_n)$ by $\hat{G}_n(sa_n)$, $0 \leqslant s \leqslant 1$, where $\hat{G}_n$ is the product-limit estimator of G_n. This estimator is obtained by formally considering the Y_{ni} as the variables censored by the X_i, in other words $\hat{G}_n$ is obtained from the expression on the right in (2.8) by replacing the $\tilde{\delta}_{ni}$ with $1 - \tilde{\delta}_{ni}$. Estimating γ is essentially the problem of estimating the extreme value index of a c.d.f. in the domain of min-attraction of a c.d.f. of Weibull type. Various choices for the estimator are possible, a particularly simple one being

$$\hat{\gamma}_n = \log \hat{F}_n(a_n)/\log a_n. \tag{3.1}$$

It is not overly hard to show that

$$\hat{\gamma}_n \to_p \gamma, \quad \text{as } n \to \infty. \tag{3.2}$$

For us $\hat{\gamma}_n$ is just *any* estimator satisfying (3.2). Finally we propose

$$\hat{D}_n(1) = \int_0^1 \frac{1}{1 - \hat{G}_n(sa_n)} \, ds^{\hat{\gamma}_n}, \tag{3.3}$$

as an estimator of $D(1)$. As before let W be a standard Wiener process and let $c = c(\alpha)$ be such that

$$\mathbb{P}\left\{ \sup_{0 \leqslant t \leqslant 1} |W(t)| \geqslant c \right\} = \alpha, \quad 0 < \alpha < 1. \tag{3.4}$$

THEOREM 3.1 Under assumptions (1.1)–(1.4) and (3.2) we have

$$\lim_{n\to\infty} \mathbb{P}\left\{\hat{F}_n(t) - c\sqrt{\frac{\hat{F}_n(a_n)\,\hat{D}_n(1)}{n}} < F(t) < \hat{F}_n(t) + c\sqrt{\frac{\hat{F}_n(a_n)\,\hat{D}_n(1)}{n}},\quad 0 \leqslant t \leqslant a_n\right\} = 1 - \alpha. \tag{3.5}$$

Proof From (2.12) it is immediate that

$$\sqrt{\frac{n}{F(a_n)}} \sup_{0\leqslant t\leqslant a_n} |\hat{F}_n(t) - F(t)| \to_d \sup_{0\leqslant t\leqslant 1} |W \circ D(t)|, \quad \text{as } n \to \infty. \tag{3.6}$$

Noticing that D is an increasing function, mapping $[0,1]$ onto $[0, D(1)]$, it follows that $\sup_{0\leqslant t\leqslant 1} |W \circ D(t)| = \sup_{0\leqslant t\leqslant D(1)} |W(t)| =_d \sqrt{D(1)} \sup_{0\leqslant t\leqslant 1} |W(t)|$ and hence (3.6) implies

$$\sqrt{\frac{n}{F(a_n)D(1)}} \sup_{0\leqslant t\leqslant a_n} |\hat{F}_n(t) - F(t)| \to_d \sup_{0\leqslant t\leqslant 1} |W(t)|, \quad \text{as } n \to \infty. \tag{3.7}$$

The theorem follows if it can be shown that

$$\frac{\hat{F}_n(a_n)}{F(a_n)} \to_p 1, \quad \text{and} \quad \frac{\hat{D}_n(1)}{D(1)} \to_p 1, \quad \text{as } n \to \infty. \tag{3.8}$$

The first of these statements is immediate from (2.12). For the second one, observe that

$$\hat{D}_n(1) - D(1) = \int_0^1 \left(\frac{1}{1 - \hat{G}_n(sa_n)} - \frac{1}{1 - G_n(sa_n)}\right) ds^{\hat{\gamma}_n} + \int_0^1 \frac{1}{1 - G(s)}\, d(s^{\hat{\gamma}_n} - s^{\gamma}). \tag{3.9}$$

The product-limit estimator $\hat{G}_n$ is very close to the empirical c.d.f. for uncensored Y_{ni} and it is easy to prove that $\sup_{0\leqslant t\leqslant a_n} |\hat{G}_n(t) - G_n(t)| \to_p 0$, as $n \to \infty$, which entails

$$\begin{aligned} &\left|\int_0^1 \left(\frac{1}{1 - \hat{G}_n(sa_n)} - \frac{1}{1 - G_n(sa_n)}\right) ds^{\hat{\gamma}_n}\right| \\ &\leqslant \int_0^1 \left|\frac{\hat{G}_n(sa_n) - G_n(sa_n)}{(1 - \hat{G}_n(sa_n))(1 - G(s))}\right| ds^{\hat{\gamma}_n} \\ &= o_p(1) \int_0^1 ds^{\hat{\gamma}_n} = o_p(1), \quad \text{as } n \to \infty. \end{aligned} \tag{3.10}$$

Consequently, to prove the second part of (3.8) it suffices to show that for arbitrary $\varepsilon > 0$ and n sufficiently large

$$\mathbb{P}\left\{\left|\int_0^1 \frac{1}{1-G(s)}\, d(s^{\hat{\gamma}_n} - s^{\gamma})\right| \geqslant \varepsilon\right\} \leqslant \varepsilon. \tag{3.11}$$

Without loss of generality we may and will assume that $\hat{\gamma}_n > 0$. Integration by parts yields

$$\left|\int_0^1 \frac{1}{1-G(s)}\, d(s^{\hat{\gamma}_n} - s^{\gamma})\right| = \int_0^1 |s^{\hat{\gamma}_n - \gamma} - 1|\, s^{\gamma}\, d\frac{1}{1-G(s)}. \tag{3.12}$$

Let us first note the simple fact that $\sup_{0<\eta\leqslant s\leqslant 1} |s^a - 1| \leqslant |\eta^a - 1|$ for each $a \in \mathbb{R}$. Now let us choose $\delta = \varepsilon/\{4\int_0^1 s^{\gamma}\, d((1-G(s))^{-1})\}$ for ε sufficiently small to ensure that $\delta \leqslant 1/2$, and let us define $\hat{\eta}_n = (1-\delta)^{1/|\hat{\gamma}_n - \gamma|}$. We have

$$\int_{\hat{\eta}_n}^1 |s^{\hat{\gamma}_n - \gamma} - 1|\, s^{\gamma}\, d\frac{1}{1-G(s)} \leqslant \int_0^1 |(1-\delta)^{(\hat{\gamma}_n - \gamma)/|\hat{\gamma}_n - \gamma|} - 1|\, s^{\gamma}\, d\frac{1}{1-G(s)}$$

$$\leqslant \int_0^1 \left\{\max_{i=\pm 1} |(1-\delta)^i - 1|\right\} s^{\gamma}\, d\frac{1}{1-G(s)} \leqslant 2\delta \int_0^1 s^{\gamma}\, d\frac{1}{1-G(s)} = \varepsilon/2. \tag{3.13}$$

Since $\hat{\eta}_n \to_p 0$, as $n \to \infty$, for any $0 < \delta < 1$ we also have with arbitrary high probability for sufficiently large n that

$$\int_0^{\hat{\eta}_n} |s^{\hat{\gamma}_n - \gamma} - 1|\, s^{\gamma}\, d\frac{1}{1-G(s)} \leqslant \int_0^{\hat{\eta}_n} d\frac{1}{1-G(s)} = \frac{G(\hat{\eta}_n)}{1-G(\hat{\eta}_n)} \leqslant \varepsilon/2. \tag{3.14}$$

This completes the proof of (3.11) and hence of (3.5). Q.E.D.

The class of c.d.f.'s F satisfying (1.3) and (1.4) contains c.d.f.'s with derivative $F'(0)$ which is either zero, finite nonzero, or infinite, partly depending on γ. Henceforth we will restrict ourselves to the subclass $\mathscr{F}$ of c.d.f.'s with the following properties: (1.3) and (1.4) are fulfilled; a continuous second derivative exists on $(0, \varepsilon]$, for some $\varepsilon > 0$; $\lim_{t\downarrow 0} f(t)$ exists ($f = F'$) and equals $f(0)$, say, where $f(0) = \infty$ is admitted; $\lim_{t\downarrow 0} f'(t)$ is finite if $f(0)$ is finite. It should be noted that finiteness of $f(0)$ entails that $F'(0) = f(0)$ so that F' is continuous on $[0, \varepsilon]$ with continuous bounded derivative f' on $(0, \varepsilon]$ in that case.

For practical purposes it is interesting to know that failure is unlikely to occur immediately. Consequently for some $0 < c < \infty$ we are interested in testing the null hypothesis H_0: $f(0) \geqslant c$ (including $f(0) = \infty$) versus the alternative H_1: $f(0) < c$ (including $f(0) = 0$). Furthermore we introduce a kernel K: $\mathbb{R} \to \mathbb{R}$ which is of bounded variation on $[0, 1]$, zero outside $[0, 1]$, and which satisfies $\int_0^1 K(t)\, dt = 1$. As a test statistic we introduce

$$\hat{f}_n(0) = \frac{1}{a_n} \int_{-\infty}^{\infty} K\left(\frac{t}{a_n}\right) d\hat{F}_n(t), \tag{3.15}$$

which is an estimator for the density f at the point zero. Write also (cf. (3.3)):

$$\hat{D}_n(t) = \int_0^t \frac{1}{1-\hat{G}_n(sa_n)}\, ds^{\hat{\gamma}_n}, \quad 0 \leqslant t \leqslant 1. \tag{3.16}$$

THEOREM 3.2 In addition to (1.1)–(1.4) and (3.2), let us assume that $na_n^3 \to 0$, as $n \to \infty$. An asymptotically size $\alpha \in (0,1)$ test for testing H_0: $f(0) \geqslant c \in (0,\infty)$ versus H_1: $f(0) < c$ is obtained when we reject H_0 if

$$\sqrt{\frac{na_n^2}{\hat{F}_n(a_n)\int_0^1 \hat{D}_n(t)\, dK^2(t)}}\,(\hat{f}_n(0) - c) \leqslant \Phi^{-1}(\alpha), \tag{3.17}$$

where Φ is the standard normal c.d.f. The test is consistent against any alternative covered by H_1.

Proof Let us introduce $f_n(0) = (1/a_n)\int_{-\infty}^{\infty} K(t/a_n)\, dF(t)$, and note that via integration by parts (2.12) entails

$$\sqrt{\frac{n}{F(a_n)}}\, a_n(\hat{f}_n(0) - f_n(0)) \to_d -\int_0^1 W \circ D(t)\, dK(t), \quad \text{as } n \to \infty, \tag{3.18}$$

since integration with respect to K is a continuous functional on $D[0,1]$. The random variable on the right in (3.18) is normal with mean 0 and variance $\int_0^1 D(t)\, dK^2(t)$. A weakly consistent estimator for this variance is given by $\int_0^1 \hat{D}_n(t)\, dK^2(t)$. Since $\sup_{0 \leqslant t \leqslant 1} |\hat{D}_n(t) - D(t)|$ is bounded by the sum of the expressions on the right in (3.10) and (3.12) it follows at once from the last part of the proof of Theorem 3.1 that $\sup_{0 \leqslant t \leqslant 1} |\hat{D}_n(t) - D(t)| \to_p 0$, as $n \to \infty$. Hence we have $|\int_0^1 \hat{D}(t)\, dK^2(t) - \int_0^1 D(t)\, dK^2(t)| \leqslant \sup_{0 \leqslant t \leqslant 1} M |\hat{D}_n(t) - D(t)| \to_p 0$, as $n \to \infty$, where M is the mass assigned to $[0,1]$ by the total variation measure determined by K^2. Jointly with the first statement in (3.8) this yields

$$\sqrt{\frac{na_n^2}{\hat{F}_n(a_n)\int_0^1 \hat{D}_n(t)\, dK^2(t)}}\,(\hat{f}_n(0) - f_n(0)) \to_d \mathcal{N}(0,1), \quad \text{as } n \to \infty. \tag{3.19}$$

Next let us replace $f_n(0)$ with c in the expression on the left in (3.19) and first show that

$$\sqrt{\frac{n}{F(a_n)}}\left\{\int_0^1 K(t)\, dF(ta_n) - ca_n\right\} \to \begin{cases} \infty, & f(0) > c \\ 0, & f(0) = c, \\ -\infty, & f(0) < c \end{cases} \quad \text{as } n \to \infty. \tag{3.20}$$

If $f(0) = c$ we have

$$\int_0^1 K(t)\, dF(ta_n) - ca_n = -\int_0^1 F(ta_n)\, dK(t) - ca_n$$

$$= -\int_0^1 (ca_n t + O(a_n^2))dK(t) - ca_n = O(a_n^2), \quad \text{as } n \to \infty. \tag{3.21}$$

Because $\sqrt{n/F(a_n)} = O(\sqrt{n/a_n})$ it follows that the expression on the left in (3.20) is of order $O(\sqrt{na_n^3}) = o(1)$, as $n \to \infty$. The remaining two cases can be dealt with in a similar manner. It is clear that (3.8), (3.19), and (3.20) imply the claims of the theorem. Q.E.D.

If the a_n are unknown one should realize that the conditions do not uniquely determine such a sequence. In fact, condition (1.1) is equivalent with the condition that, for any number $0 < q < 1$, there exists a continuous c.d.f. G with $G(1) = q$ such that $G_n(t) = G(t/a_n)$ $0 \leqslant t \leqslant a_n$, for some sequence of strictly positive numbers $(a_n)_{n \in \mathbb{N}}$ satisfying (1.2) To see that (1.1) implies this condition choose $c > 0$ in such a way that $G(c) = q$, and observe that $\mathbb{P}\{Y_{ni}/ca_n \leqslant t\} = \mathbb{P}\{Y_{ni}/a_n \leqslant ct\} = G(ct)$, $0 \leqslant t \leqslant 1$. Hence the condition is fulfilled for the c.d.f. $\tilde{G}(\cdot) = G(c\cdot)$ with $\tilde{G}(1) = q$, indeed, and the sequence $\tilde{a}_n = ca_n$. This sequence also satisfies (1.2), as easily follows from (1.3) and (1.4).

Hence, let us assume that $G(1) = q$ for a given $0 < q < 1$ that we are free to choose. Consequently we have $G_n(a_n) = q$, $a_n = G_n^{-1}(q)$, which suggests

$$\hat{a}_n = \hat{G}_n^{-1}(q), \tag{3.22}$$

as a suitable estimator of a_n. Apparently the estimator of a_n is a quantile estimator and for proper dealing with such an estimator it is not surprising that an additional smoothness condition on G_n will be needed. The relation between G_n and G leads to the assumption

$$G'(1) \text{ exists} \quad \text{and} \quad G'(1) > 0, \tag{3.23}$$

under which the key result

$$\frac{\hat{a}_n}{a_n} \to_p 1, \quad \text{as } n \to \infty, \tag{3.24}$$

can be established.

Let us give an informal proof of this claim and first observe that the original censoring of the Y_{ni} by the X_i is equivalent with the censoring of the $\tilde{Y}_{ni} = Y_{ni}/a_n$ by the $\tilde{X}_{ni} = X_i/a_n$, where the $\tilde{Y}_{ni}$ are i.i.d, with c.d.f. G as has been observed below (1.2). The censoring of the $\tilde{Y}_{ni}$ by the $\tilde{X}_{ni}$ now is "light" so that the corresponding product-limit estimator which is, of course, equal to $\hat{G}_n(a_n\cdot)$ satisfies the usual properties like the Glivenko–Cantelli convergence

$$\sup_{0 \leqslant t \leqslant c} |\hat{G}_n(a_n t) - G(t)| \underset{a.s.}{\to} 0, \quad \text{as } n \to \infty, \tag{3.25}$$

for any $1 < c < G^{-1}(1)$, in particular. The inverse of $\hat{G}_n(a_n\cdot)$ being $\hat{G}_n^{-1}(\cdot)/a_n$, this property entails as usual the convergence of the quantiles

$$\hat{G}_n^{-1}(q)/a_n = \hat{a}_n/a_n \to_p G^{-1}(q) = 1, \quad \text{as } n \to \infty, \tag{3.26}$$

provided that (3.23) is satisfied. This settles (3.24), with the help of which it is rather straightforward to prove the following result. The details will be omitted.

THEOREM 3.3 Under the assumptions of Theorem 3.1 and Theorem 3.2 and (3.23) or (3.24) the results of these theorems remain true with a_n consistently replaced by $\hat{a}_n$.

4. SOME REMARKS

4.1. Tail empirical processes are of theoretical interest in their own right and the same could be said about the present generalization to the censored case. Observe that the definition of G_n (in (1.1)) and the window $[0, a_n]$ are "balanced" in such a way that exactly G shows up in the time transformation D of the limiting Wiener process. E.g., choosing a censoring distribution not depending on n, but keeping the same window, would lead to a limiting process in which the censoring distribution plays no role.

4.2. Theorem 2.1 can be extended to the case where the processes ξ_n, β_n and $W \circ D$ are divided by a weight function; also functional laws of the iterated logarithm for ξ_n and β_n are readily derived similarly, cf. Einmahl (1992).

4.3. An admissible choice for the kernel K in (3.15) is the indicator function $\mathbb{1}_{[0,1]}$ in which case $\hat{f}_n(0)$ reduces to $\hat{F}_n(a_n)/a_n$.

4.4. An alternative estimator of $D(t)$ in the variance $\int_0^1 D(t)dK^2(t)$ of the random variable on the right in (3.18) is given by $\hat{\hat{D}}_n(t) = \int_0^t (1 - \hat{G}_n(sa_n))^{-1} ds$, since $\gamma = 1$ under H_0.

4.5. The condition $na_n^3 \to 0$, as $n \to \infty$, in Theorem 3.2 is in fact a restriction on the model. If the condition is not fulfilled a result like the one in Theorem 3.2 does not exist. Further smoothness of F, however, can be used to relax this restriction.

4.6. Local confidence bands for Λ and tests on $\lambda(0) = \Lambda'(0)$ follow along similar lines from (2.13), but since $\lambda(0) = f(0)$, Theorem 3.2 can also be used directly for testing on $\lambda(0)$.

Acknowledgement

Both authors were partially supported by NATO-grant CRG 910397.

References

[1] Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship, *Ann. Statist.* **2**, 437–453.
[2] Einmahl, J. H. J. (1992). The a.s. behavior of the weighted empirical process and the LIL for the weighted tail empirical process, *Ann. Statist.* **20**, 681–695.
[3] Einmahl, J. H. J. and Koning, A. J. (1992). Limit theorems for a general weighted process under random censoring, *Canadian J. Statist.* **20**, 77–89.
[4] Gill, R. D. and Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis, *Ann. Statist.* **18**, 1501–1555.

[5] Shorack, G. R. and Wellner, J. A. *Empirical Processes with Applications to Statistics*, Wiley, New York, 1986.

[6] Wellner, J. A. (1985). A heavy censoring limit theorem for the product limit estimator, *Ann. Statist.* **13**, 150–162.

APPENDIX

Proof of Theorem 2.1

Let us first prove (2.13). It is immediate from Shorack & Wellner (1986, formula (16) of Section 7.1, Theorem 1 of Section 7.2), and (2.5) that β_n can be rewritten as

$$\beta_n(t) = \frac{U_n^*(ta_n)}{\sqrt{F(a_n)}(1-H_n^-(ta_n))} - \int_0^{ta_n} \frac{U_n^*(s)}{\sqrt{F(a_n)}} d\frac{1}{1-\hat{H}_n^-(s)} + \int_0^{ta_n} \frac{U_n^-(s)}{\sqrt{F(a_n)}(1-\hat{H}_n^-(s))} d\Lambda(s), \quad 0 \leqslant t \leqslant 1, \tag{A.1}$$

provided that $a_n \leqslant \tilde{Z}_{nn}$. The obvious relation $H_n = 1-(1-F)(1-G_n)$ and (1.1) and (1.2) entail that $H_n(t) = G(t/a_n) + F(t)(1-G(t/a_n))$ and hence that $H_n(ta_n) = G(t) + F(ta_n)(1-G(t)) \to G(t)$, $0 \leqslant t \leqslant 1$, as $n \to \infty$. Hence it follows from standard empirical process theory that $\{U_n^-(ta_n), 0 \leqslant t \leqslant 1\}$ converges weakly to $\{B \circ G(t), 0 \leqslant t \leqslant 1\}$ on $D[0,1]$, as $n \to \infty$, where B is a standard Brownian bridge. This yields

$$\begin{aligned}
&\sup_{0\leqslant t\leqslant 1} \left| \int_0^{ta_n} \frac{U_n^-(s)}{\sqrt{F(a_n)}(1-\hat{H}_n^-(s))} d\Lambda(s) \right| \\
&= \sup_{0\leqslant t\leqslant 1} \left| \int_0^t \frac{U_n^-(sa_n)}{1-\hat{H}_n^-(sa_n)} d\frac{\Lambda(sa_n)}{\sqrt{F(a_n)}} \right| \\
&= \frac{\Lambda(a_n)}{\sqrt{F(a_n)}} O_p(1) = \sqrt{F(a_n)}\, O_p(1), = o_p(1), \quad \text{as } n \to \infty,
\end{aligned} \tag{A.2}$$

employing (2.4).

So we can focus on the first two terms on the right in (A.1). Another well-known relation, $H_n^*(t) = \int_0^t (1-G_n(s))dF(s)$, $0 \leqslant t \leqslant 1$, yields $H_n^*(ta_n)/F(a_n) \to \int_0^t (1-G(s))\, ds^\gamma = \tilde{D}(t)$, $0 \leqslant t \leqslant 1$, as $n \to \infty$. It has been shown in, e.g., Einmahl and Koning (1992) that $U_n^*(ta_n)$ behaves like an ordinary empirical process based on n i.i.d. observations from the c.d.f. $H_n^*(ta_n)$, which entails that $U_n^*(ta_n)/\sqrt{F(a_n)}$ has variance $H_n^*(ta_n)\,(1-H_n^*(ta_n))/F(a_n) \to \tilde{D}(t)$, as $n \to \infty$. This explains that there exists a special construction such that

$$\sup_{0\leqslant t\leqslant 1} \left| \frac{U_n^*(ta_n)}{\sqrt{F(a_n)}} - \tilde{W} \circ \tilde{D}(t) \right| \to_p 0, \quad \text{as } n \to \infty, \tag{A.3}$$

where $\tilde{W}$ is a standard Wiener process. This implies

$$\sup_{0\leq t\leq 1}\left|\frac{U_n^*(ta_n)}{\sqrt{F(a_n)}(1-\hat{H}_n^-(ta_n))}-\frac{\tilde{W}\circ\tilde{D}(t)}{1-G(t)}\right|\to_p 0, \quad \text{as } n\to\infty, \tag{A.4}$$

because

$$\sup_{0\leq t\leq 1}|\hat{H}_n^-(ta_n)-G(t)|\to_p 0, \quad \text{as } n\to\infty. \tag{A.5}$$

Using (A.3) again we see that

$$\sup_{0\leq t\leq 1}\left|\int_0^{ta_n}\frac{U_n^*(s)}{\sqrt{F(a_n)}}\,d\frac{1}{1-\hat{H}_n^-(s)}-\int_0^t \tilde{W}\circ\tilde{D}(s)\,d\frac{1}{1-\hat{H}_n^-(sa_n)}\right|\to_p 0, \quad \text{as } n\to\infty. \tag{A.6}$$

Moreover, subtle application of the Helly–Bray theorem (cf. Shorack and Wellner (1986, p. 309)) yields

$$\sup_{0\leq t\leq 1}\left|\int_0^t \tilde{W}\circ\tilde{D}(s)\,d\left(\frac{1}{1-\hat{H}_n^-(sa_n)}-\frac{1}{1-G(s)}\right)\right|\to_p 0, \quad \text{as } n\to\infty. \tag{A.7}$$

Combining (A.1), (A.2), (A.4), (A.6), and (A.7) we obtain

$$\sup_{0\leq t\leq 1}\left|\beta_n(t)-\left(\frac{\tilde{W}\circ\tilde{D}(t)}{1-G(t)}-\int_0^t \tilde{W}\circ\tilde{D}(s)\,d\frac{1}{1-G(s)}\right)\right|\to_p 0, \quad \text{as } n\to\infty. \tag{A.8}$$

Finally, routine considerations show that the process

$$W\circ D(t)=\frac{\tilde{W}\circ\tilde{D}(t)}{1-G(t)}-\int_0^t \tilde{W}\circ\tilde{D}(s)\,d\frac{1}{1-G(s)}, \quad 0\leq t\leq 1, \tag{A.9}$$

is a zero mean Wiener process with covariance function $D(s)\wedge D(t)$ for $s,t\in[0,1]$ which entails (2.13) using (A.8).

Next let us consider (2.12). We start with the well-known identity

$$\xi_n(t)=(1-F(ta_n))\int_0^t\frac{1-\hat{F}_n^-(sa_n)}{1-F(sa_n)}\,d\beta_n(s), \quad 0\leq t\leq 1; \tag{A.10}$$

see, e.g., Shorack and Wellner (1986, Proposition 1 of Section 7.2). Using (1.2) and the result (2.13) that we just proved, it suffices to show that

$$\sup_{0\leq t\leq 1}\left|\int_0^t\frac{1-\hat{F}_n^-(sa_n)}{1-F(sa_n)}\,d\beta_n(s)-\beta_n(t)\right|\to_p 0, \quad \text{as } n\to\infty. \tag{A.11}$$

This is equivalent with showing that

$$\sup_{0 \leqslant t \leqslant 1} \left| \int_0^t \frac{F(sa_n) - \hat{F}_n^-(sa_n)}{1 - F(sa_n)} d\beta_n(s) \right| \to_p 0, \quad \text{as } n \to \infty. \tag{A.12}$$

Integration by parts shows that the expression on the left in (A.12) is bounded by

$$\begin{aligned} &\sup_{0 \leqslant t \leqslant 1} \left| \frac{F(ta_n) - \hat{F}_n^-(ta_n)}{1 - F(ta_n)} \right| |\beta_n(t)| \\ &+ \sup_{0 \leqslant t \leqslant 1} \int_0^t |\beta_n(s)| d \frac{\hat{F}_n^-(sa_n)}{1 - F(sa_n)} \\ &+ \sup_{0 \leqslant t \leqslant 1} \int_0^t |\beta_n(s)| d \frac{F(sa_n)}{1 - F(sa_n)}. \end{aligned} \tag{A.13}$$

First let us note that

$$\sup_{0 \leqslant t \leqslant 1} |\hat{F}_n(ta_n) - F(ta_n)| \to_p 0, \quad \text{as } n \to \infty. \tag{A.14}$$

To see this observe that $\hat{F}_n(ta_n) - F(ta_n)$ equals the expression on the right in (A.10) with $\beta_n(t)$ replaced by $\beta_n^*(t) = \sqrt{\frac{F(a_n)}{n}}\,\beta_n(t)$. From (2.13) we obtain that the three terms in (A.13) with β_n replaced by β_n^* converge to 0 in probability, so that (A.11) holds true with β_n^* instead of β_n, and hence (A.14) follows.

Now (A.12), and hence (2.12), easily follows by combining (A.14) and the facts that $\sup_{0 \leqslant t \leqslant 1} |\beta_n(t)| = O_p(1)$, and $F(a_n) \to 0$, as $n \to \infty$. Q.E.D.